COMM 599: Applied Quantitative Methods.
4 Units

Fall 2024 – Monday - 6:30-9:30pm
Location: ASC 331

Instructor: Joshua Clark
Office: Virtual
Office Hours: Friday Afternoon 1-3pm.
Contact info: joshuaac@usc.edu. Booking Link https://calendly.com/josh-aclark/office-hours

Course Description

This course is meant to prepare students for working in a modern data and analytics team and communicate with peers and stakeholders. The goal is to allow students to build a project and portfolio, which they can show to prospective employers in order to demonstrate the skills that they have acquired thus far and communicate their value to a company. This course requires a basic familiarity with data science techniques and focuses on how to explain and communicate complex statistical concepts to both collaborators and a non-technical audience.

Student Learning Outcomes

Throughout this course students will learn how to.
- Collect a unique dataset distinctive to their interests.
- Manipulate this data to deal with errors, missing values and other issues.
- Collaborate like a data science team, using common co-working methodologies employed in industry.
- Learn to document their own code.
- Undergo a data peer review process and integrate changes into a codebase.
- Create a final portfolio of code, data and visualizations which can be paired with a resume.

Prerequisite(s): There are no course requirements for this class, but students are expected to have a familiarity with Python or R and basic data science methods.
Recommended Preparation: An active Github account is required, please create one before class.

Course Notes

In mimicking a data organization, this course will use a private Github repository for the provision and submission of assignments. Students will be expected to have their own Github account and will leave the course with a repository that can be used as a portfolio. All lecture materials will also be made available via the course git repository.

Because this course mimics a Data Organization Students are required to have a working familiarity of a statistical programming language. Python is preferred but R is acceptable. Students who do not have this level of familiarity will face significant challenges given the technical and hands on nature of the course.

Required Readings and Supplementary Materials

There is no required textbook for this class. All readings are linked in the course list below.

Optional Readings and Supplementary Materials

As a supplemented reference, I would recommend getting An Intro to Statistical Learning, available here.

Technological Proficiency and Hardware/Software Required

This course requires access to a computer running Windows, MACOS or Linux. Laptops are available through the USC Computing Center Laptop Loaner Program. Additionally, we will be using a number of free and open source tools including but not limited to Python.

Description and Assessment of Assignments

The core element of this class is a portfolio project. Students are expected to find a data source, work with their peers to identify a transformation or data product derived from this data source that creates value and then execute this plan.

Pitch Deck/Summary - A brief presentation describing the data source and problem supporting a student's project. Given early in the semester, the pitch deck should be a short, high-level presentation meant to convince the audience that this project is a good use of time and resources. It should be accompanied by a one-page cover sheet summarizing the pitch.

Weekly Stand Up - Following practice of data science teams, students are expected to give a short verbal summary of their progress, challenges and achievements every week. If unable to make a class, a written summary can serve as a replacement.

Peer Review - As students make progress on their project, they will commit code to the class's common repository. Each student will have a designated reviewer who is required to review and approve commits to the repository. Informative and timely reviews will earn credit for the reviewer.

FInal Project Assessment - Students' final project will be assessed on two fronts:
- Quality of data product - Points will be assigned based on the scope and quality of the submitted code and data product. Work will be assessed on ease of use, readability of code, usefulness of final project and overall presentation/documentation.
- Final Presentation - Students will be expected to give a 15-20 minute presentation describing their work, the challenges and the value that their data product delivers. Presentations should assume a non-technical audience with little familiarity with the project and demonstrate the value created by the work in a clear, concise and non-confusing way. Presentations will be spread over two weeks to allow time for questions and dialogue, students presenting on the second week will have a higher expectation of polish in their work to account for the extra time.
- Final code review, the final portfolio is open to an entire class code review.

Participation

Participation will be assessed through the stand-up meetings. Students can make up missed standups by providing written summaries of their reports, however, attendance is strongly encouraged to benefit from instruction and peer review within the classroom setting.

Grading Breakdown

Description of assessments and corresponding points and percentage of grade.

Assessment Tool (assignments)	Points	% of Grade
Pitch Deck	15	15%
Stand Up Updates	15	15%
Peer Review	15	15%

Assessment Tool (assignments)	Points	% of Grade
Final Code Review	25	25%
Final Presentation	30	30%
TOTAL		**100%**

Course Grading Scale

Letter grade and corresponding numerical point range		
94% to 100%: A	80% to 83%: B- (B minus)	67% to 69%: D+ (D plus)
90% to 93%: A- (A minus)	77% to 79%: C+ (C plus)	64% to 66%: D
87% to 89%: B+ (B plus)	74% to 76%: C	60% to 63%: D- (D minus)
84% to 86%: B	70% to 73%: C- (C minus)	0% to 59%: F

Grading Standards
What each letter grade demonstrates.

Letter Grade	Description
A	**Solid contributions, elevates their peers and produces good quality code. Can communicate complex data processes to technical and stakeholder audiences.**
B	**Good work, produces effective code and can communicate simple data processes. Work may lack scope, or there may be room for improvement in tackling more complex topics.**
C	**Satisfactory/Fair; reasonable knowledge and understanding of subject matter. Demonstrates basic understanding of data science principles but has not produced or communicated results effectively.**
D	**Marginal; minimal knowledge and understanding of subject matter; more than one significant shortcoming. Either serious issues with the final portfolio or systemic problems with the communication of data topics.**
F	**Failing; unacceptably low level of knowledge and understanding of subject matter; deficiencies indicate lack of understanding.**

Grading Timeline
All grades will be submitted within a week of receiving the assignment. Feedback will be provided through email and comments on the github repo.

Course Policies
You are welcome/expected to use generative AI tools (e.g. ChatGPT, Dall-e, etc.) in this class as doing so aligns with the course learning goal of developing code for a portfolio in a way that mimics a member of a data team.. You are responsible for the information you submit based on an AI query (for instance, that it does not violate intellectual

property laws, or contain misinformation or unethical content). Your use of AI tools must be properly documented and cited in order to stay within university policies on academic integrity.

Copying of code from Chat-GPT, Stackoverflow and other sources is allowed. Any code blocks taken from these sources should have a comment or link to the source to help with reproducibility.

Assignment Submission Policy

Assignments are due by the start of class. Grades for the stand-up portion of the class will be determined by engagement within the classroom or submission of a written summary of your progress if you have to miss a class.

Attendance

If students are unable to attend and cannot provide an update in the Stand-Up Meeting then they can provide a written update to the entire class on their progress and blockers for the past week.

Classroom Norms

The classroom will be run like an industry data team, students are expected to maintain a level of professionalism and respect that would mimic that in a workplace. Additionally students are collaborating on their code reviews and projects and should strive to be responsive and helpful to their peers.

As with any software development process students are expected to maintain security. Do not share access keys, passwords, code which is not yours or other secrets with third parties without approval from the instructor.

Communication

Please do not hesitate to reach out if you cannot attend office hours or need extra help with assignments. The best route of contact is by email. Barring extenuating circumstances, emailed requests will be responded to within 24 hours.

Please use the Calendly link to book office hours.

Course Schedule

Important note to students: Be advised that this calendar is subject to change based on the progress of the class, news events, or guest speaker availability.

	Topics/Daily Activities	Readings and Homework	Deliverable/Due Dates
Week 1 **Dates: 8/26-8/30**	Introduction to class, discussion of project deliverables, tech setup and bug hunting.		
Week 2 **Dates: 9/2-9/6**	Holiday		[**Labor Day**: Monday, September 2]
Week 3 **Dates: 9/9-9/13**	Understanding the modern data team.	Harvard Business Review: Data Scientist.	Special Stand Up - Describe your career goals within the data space.

	Topics/Daily Activities	Readings and Homework	Deliverable/Due Dates
	Where and how to find data.	The Ambiguity of Data Science Team Roles and the Need for a Data Science Workforce Framework List of open data sources Los Angeles Open Data	
Week 4 **Dates: 9/16-9/20**	What is a data product? Finding data, work scoping, timelines and deliverables	Tidy Data Map Reduce Simplified Data Processing on Large Clusters State of Data Science W3 SQL Reference (if you need it)	Stand Up - data exploration
Week 5 **Dates: 9/23-9/27**	Communication within data teams. Peer review and project management.	The Agile Manifesto Agile Data Science Hello World - Github Git and Github for beginners.	Pitch Deck presentations
Week 6 **Dates: 9/30-10/4**	Data visualization, dashboarding and avoiding common graphing mistakes.	Student reading assignment, find and share favorite data visualization from media or culture. The Grammar of Graphics	Stand Up - present visualization as part of stand up.
Week 7 **Dates:** **10/7-10/11**	Experiments and communication. AB Testing and data driven product development	Avoiding the Pitfalls of AB Testing Here's How Cornell Scientist Brian Wansink Turned Shoddy Data Into Viral Studies About How We Eat	Stand Up

	Topics/Daily Activities	Readings and Homework	Deliverable/Due Dates
		Bayesian Methods for Hackers, Chp 1 & 2, PYMC Version (Optional)	
Week 8 Dates: 10/14-10/18	Delivering supervised learning results. Classification, regression and understanding models.	SHAP Scoring Repo and Documentation "Why Should I Trust You?" Explaining the Predictions of Any Classifier Interpretable ML, Chapters 1-3	Stand Up - midterm course evaluation survey.
Week 9 Dates: 10/21-10/25	Clustering and unsupervised models.	Clustering in machine learning Comprehensive Survey of Clustering Clustering, That's Cute.	Stand Up - code should have initial peer review by this point at the latest.
Week 10 Dates: 10/28-11/1	Machine learning, neural networks, LLMs.	Statistical Learning, the two Cultures. Stop explaining black box machine learning models for high stakes decisions Sparks of AGI (skim)	Stand Up
Week 11 Dates: 11/4-11/8	Working with qualitative and mixed methods researchers.	When User Experience Designers Partner with Data Scientists Investigating How Experienced UX Designers Effectively Work with Machine Learning	Stand Up
Week 12 Dates: 11/11-11/15	Homestretch hack-a thon week. Prep for stakeholder presentations.		

	Topics/Daily Activities	Readings and Homework	Deliverable/Due Dates
Week 13 **Dates:** **11/18-11/22**	Resume and CV Workshop	Intro to LaTeX Overleaf - LaTeX in 30 Github pages.	Bring your current CV/Resume.
Week 14 **Dates:** **11/25-11/29**	Stakeholder Presentations pt1	Reading summaries, code and decks from other student's projects for feedback and questions. Final Code review, entire class comments on code.	Presentation decks and errata due for early presenters.
Week 15 **Dates: 12/2-12/6**	Stakeholder Presentations pt2	Reading summaries, code and decks from other student's projects for feedback and questions. Final Code review, entire class comments on code.	Presentation decks and errata due. Last chance of peer review of code
STUDY DAYS **Dates:** **12/7-12/10**			
FINAL EXAM **PERIOD** **Date:** **12/16**	Portfolio due		Complete portfolio (resume, code, deck and errata) need to be committed to github by 9pm at the latest.

Statement on Academic Conduct and Support Systems

Academic Integrity
The University of Southern California is foremost a learning community committed to fostering successful scholars and researchers dedicated to the pursuit of knowledge and the transmission of ideas. Academic misconduct is in contrast to the university's mission to educate students through a broad array of first-rank academic, professional, and extracurricular programs and includes any act of dishonesty in the submission of academic work (either in draft or final form).

This course will follow the expectations for academic integrity as stated in the USC Student Handbook. All students are expected to submit assignments that are original work and prepared specifically for the course/section in this academic term. You may not submit work written by others or

"recycle" work prepared for other courses without obtaining written permission from the instructor(s). Students suspected of engaging in academic misconduct will be reported to the Office of Academic Integrity.

Other violations of academic misconduct include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

Academic dishonesty has a far-reaching impact and is considered a serious offense against the university. Violations will result in a grade penalty, such as a failing grade on the assignment or in the course, and disciplinary action from the university itself, such as suspension or even expulsion.

For more information about academic integrity see the student handbook or the Office of Academic Integrity's website, and university policies on Research and Scholarship Misconduct.

Please ask your instructor if you are unsure what constitutes unauthorized assistance on an exam or assignment or what information requires citation and/or attribution.

AI Generators Policy
Students are free to use Generative AI as a code-writing copilot and for hunting bugs. If the AI Agent is generating entire chunks of code then the prompt should be cited with a link the the prompt and discussion embedded in the code, or a copy of the prompt if deep linking to the conversation is not available.

Course Content Distribution and Synchronous Session Recordings Policies
USC has policies that prohibit recording and distribution of any synchronous and asynchronous course content outside of the learning environment.

Recording a university class without the express permission of the instructor and announcement to the class, or unless conducted pursuant to an Office of Student Accessibility Services (OSAS) accommodation, is prohibited. Recording can inhibit free discussion in the future, and thus infringe on the academic freedom of other students as well as the instructor. (Living our Unifying Values: The USC Student Handbook, page 13).

Distribution or use of notes, recordings, exams, or other intellectual property, based on university classes or lectures without the express permission of the instructor for purposes other than individual or group study. This includes but is not limited to providing materials for distribution by services publishing course materials. This restriction on unauthorized use also applies to all information, which had been distributed to students or in any way had been displayed for use in relation to the class, whether obtained in class, via email, on the internet, or via any other media. Distributing course material without the instructor's permission will be presumed to be an intentional act to facilitate or enable academic dishonestly and is strictly prohibited. (Living our Unifying Values: The USC Student Handbook, page 13).

Statement on University Academic and Support Systems

Students and Disability Accommodations:
USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who

encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

Student Financial Aid and Satisfactory Academic Progress:
To be eligible for certain kinds of financial aid, students are required to maintain Satisfactory Academic Progress (SAP) toward their degree objectives. Visit the Financial Aid Office webpage for undergraduate- and graduate-level SAP eligibility requirements and the appeals process.

Support Systems:
Annenberg Student Success Fund
The Annenberg Student Success Fund is a donor-funded financial aid account available to USC Annenberg undergraduate and graduate students for non-tuition expenses related to extra- and co-curricular programs and opportunities.

Annenberg Student Emergency Aid Fund
Awards are distributed to students experiencing unforeseen circumstances and emergencies impacting their ability to pay tuition or cover everyday living expenses. These awards are not intended to cover full-tuition expenses, but rather serve as bridge funding to guarantee students' continued enrollment at USC until other resources, such as scholarships or loans, become available. Students are encouraged to provide as much information in their application, as well as contact their academic advisor directly with questions about additional resources available to them.

Counseling and Mental Health - *(213) 740-9355 – 24/7 on call*
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

988 Suicide and Crisis Lifeline - *988 for both calls and text messages – 24/7 on call*
The 988 Suicide and Crisis Lifeline (formerly known as the National Suicide Prevention Lifeline) provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week, across the United States. The Lifeline consists of a national network of over 200 local crisis centers, combining custom local care and resources with national standards and best practices. The new, shorter phone number makes it easier for people to remember and access mental health crisis services (though the previous 1 (800) 273-8255 number will continue to function indefinitely) and represents a continued commitment to those in crisis.

Relationship and Sexual Violence Prevention Services (RSVP) - *(213) 740-9355(WELL) – 24/7 on call*
Free and confidential therapy services, workshops, and training for situations related to gender- and power-based harm (including sexual assault, intimate partner violence, and stalking).

Office for Equity, Equal Opportunity, and Title IX (EEO-TIX) - *(213) 740-5086*
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - *(213) 740-2500*
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

The Office of Student Accessibility Services (OSAS) - *(213) 740-0776*
OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

USC Campus Support and Intervention - *(213) 740-0411*
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity, Equity and Inclusion - *(213) 740-2101*
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - *UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call*
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - *UPC: (213) 740-6000, HSC: (323) 442-1200 – 24/7 on call*
Non-emergency assistance or information.

Office of the Ombuds - *(213) 821-9556 (UPC) / (323-442-0382 (HSC)*
A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

Occupational Therapy Faculty Practice - *(323) 442-2850 or* otfp@med.usc.edu
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.

About Your Instructor

Joshua (Josh) Clark (USC PhD 2016) is a gaming industry data scientist. Josh has eight years experience working at companies such as Riot Games, Meta and Jam City Games working at scale on major intellectual properties.